

Hierarchical Bayesian inference of galaxy redshift distributions from photometric surveys

Boris Leistedt,^{1,2★} Daniel J. Mortlock^{3,4} and Hiranya V. Peiris²

¹Center for Cosmology and Particle Physics, Department of Physics, New York University, New York, NY 10003, USA

²Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK

³Astrophysics Group, Imperial College London, Blackett Laboratory, Prince Consort Road, London SW7 2AZ, UK

⁴Department of Mathematics, Imperial College London, London SW7 2AZ, UK

Accepted 2016 May 27. Received 2016 May 26; in original form 2016 February 18

ABSTRACT

Accurately characterizing the redshift distributions of galaxies is essential for analysing deep photometric surveys and testing cosmological models. We present a technique to simultaneously infer redshift distributions and individual redshifts from photometric galaxy catalogues. Our model constructs a piecewise constant representation (effectively a histogram) of the distribution of galaxy types and redshifts, the parameters of which are efficiently inferred from noisy photometric flux measurements. This approach can be seen as a generalization of template-fitting photometric redshift methods and relies on a library of spectral templates to relate the photometric fluxes of individual galaxies to their redshifts. We illustrate this technique on simulated galaxy survey data, and demonstrate that it delivers correct posterior distributions on the underlying type and redshift distributions, as well as on the individual types and redshifts of galaxies. We show that even with uninformative priors, large photometric errors and parameter degeneracies, the redshift and type distributions can be recovered robustly thanks to the hierarchical nature of the model, which is not possible with common photometric redshift estimation techniques. As a result, redshift uncertainties can be fully propagated in cosmological analyses for the first time, fulfilling an essential requirement for the current and future generations of surveys.

Key words: galaxies: distances and redshifts – galaxies: statistics – cosmology: observations – large-scale structure of Universe.

1 INTRODUCTION

Testing cosmological models using the distribution of galaxies has become a routine operation thanks to large galaxy surveys such as the Sloan Digital Sky Survey (SDSS; Gunn et al. 2006) and the Canada–France–Hawaii Telescope Lensing Survey (CFHTLenS; Heymans et al. 2012). Ongoing and upcoming imaging surveys, for example the Dark Energy Survey (DES, 2012–; Abbott et al. 2005) and the Large Synoptic Survey Telescope (LSST, 2020–; Abell et al. 2009) will probe even larger volumes and will allow us to constrain the large-scale properties of the Universe at unprecedented accuracy. This will prove essential for testing our understanding of gravity, dark matter and dark energy, as well as for searching for new physics in unexplored regimes (e.g. Albrecht et al. 2006; Peacock et al. 2006; Weinberg et al. 2013).

The redshift distributions of galaxies are essential ingredients for exploiting photometric survey data. They are needed to confront measurements of the clustering and cosmic shear of galaxies with theoretical predictions. The accuracy of these distributions is

critical since any mischaracterization can translate into significant biases in the cosmological parameters inferred from data (see e.g. Cunha et al. 2012; Hildebrandt et al. 2012; Benjamin et al. 2013; Huterer, Cunha & Fang 2013; Bonnett et al. 2015). Currently, all methods for estimating redshift distributions rely on the availability of redshift estimates for all galaxies in the sample of interest. This is problematic because such estimates can suffer from systematic errors or increased sensitivity to prior information and training data, which significantly affects the quality of the redshift distribution estimates (e.g. Hildebrandt et al. 2012; Sanchez et al. 2014). The method proposed in this paper addresses these issues by simultaneously inferring the redshift distributions as well as individual redshift estimates.

Three broad classes of methods are available to estimate redshifts from noisy photometric fluxes or magnitudes: template-fitting, machine-learning, and clustering methods. Template-fitting methods (e.g. Benitez 2000; Feldmann et al. 2006; Brammer, van Dokkum & Coppi 2008) assume that each galaxy belongs to a type whose rest-frame luminosity density (simply called ‘spectral energy density’ or ‘spectrum’ below) is known. It can be taken from a library of spectral templates constructed from data

* E-mail: boris.leistedt@nyu.edu

or simulations. Redshift estimates are then obtained by predicting photometric fluxes as a function of redshift and comparing them with the observed fluxes. The second widespread class of methods involves fitting the flux–redshift relation directly using a flexible model, via machine learning techniques such as neural networks or decision trees (e.g. Collister & Lahav 2004; Kind & Brunner 2013; Sadeh, Abdalla & Lahav 2015). The third approach, sometimes referred to as ‘clustering redshifts’, delivers redshift estimates by exploiting three-dimensional spatial information, e.g. via reconstruction of the three-dimensional density field or cross-correlations of various galaxy catalogues (e.g. Matthews & Newman 2010; Jasche & Wandelt 2012; Choi et al. 2015). The lack of training and validation data at high redshift and faint magnitudes (where most of the cosmological information is) significantly hinders these three approaches. In addition, photometric calibration and distortions due to instrumental effects or observing conditions affect the quality of the measured fluxes, and hence the redshift estimates (e.g. Leistedt et al. 2015). For these reasons, photometric redshift estimation is a major challenge in the exploitation of photometric surveys. Recent comparison of methods and requirements for modern surveys can be found in e.g. Hildebrandt et al. (2010), Cunha et al. (2012), Newman et al. (2013), Sanchez et al. (2014), Abate, Newman & Schmidt (2014), Schmidt, Newman & Abate (2014).

Recent surveys such as CFHTLenS and DES obtained redshift distributions by stacking the individual redshift posterior distributions obtained with some of the methods described above (Hildebrandt et al. 2012; Bonnett et al. 2015). Despite its simplicity, this approach does not yield uncertainties on the redshift distribution estimates. Nevertheless, it is becoming increasingly important to propagate redshift uncertainties in cosmological analyses (as in the recent analysis of DES Science Verification data, Abbott et al. 2015). In this paper, we present a technique for inferring redshift distributions through a hierarchical Bayesian model which provides full posterior distributions on the redshift distributions as well as on the individual redshift estimates. This approach can be seen as a generalization of template-fitting methods to the estimation of redshift distributions. We use the library of templates from Coleman, Wu & Weedman (1980) and the type-redshift-magnitude likelihood function implemented in the BPZ method (Benitez 2000). BPZ was one of the main codes used to produce redshift distributions estimates through stacking of the individual redshifts for CFHTLenS and DES (Hildebrandt et al. 2012; Bonnett et al. 2015). As highlighted above, the common stacking approach is highly sensitive to the priors on the redshift and type distributions, which are usually assumed to have restrictive analytical forms and calibrated on training data shallower than the full samples of interest. Our model relaxes these assumptions in several ways since it infers redshift distributions from the full data set under consideration using a hierarchical probabilistic model. The posterior uncertainties in the individual redshift estimates are shrunk thanks to the addition of distribution information, and this shrinkage also includes the uncertainties in these distributions.

We describe our model for the redshift distributions in Section 2, and describe its mapping on to photometric survey data in Section 3. We extend our formalism to support arbitrary selection effects, including tomographic redshift bins, in Section 4. The framework is demonstrated on simulations in Section 5. We conclude in Section 6.

2 POPULATION MODEL

Let us consider a catalogue of galaxies with intrinsic properties type t , redshift z , and apparent magnitude m in some reference

photometric band. The properties of these galaxies are assumed to be drawn from a unit-normalized distribution $p(t, z, m | \text{galaxy, survey})$, where the form of the distribution is explicitly dependent on the abundance of the sources as a function of redshift and intrinsic properties (as determined by e.g. their luminosity function and the comoving volume per unit redshift), and the characteristics and selection effects of the survey under consideration. For concision we adopt the more compact notation $p(t, z, m)$ for this distribution, but it is important to keep in mind that it depends on particularly survey-specific effects, a point which is explored further in Section 4.

Rather than attempting to model either the galaxy population or the observational selection effects in detail, we adopt a piecewise constant representation of $p(t, z, m)$, parametrized by a set of coefficients $\{f_{ijk}\}$, such that

$$p(t, z, m | \{f_{ijk}\}) = \sum_{ijk} \frac{f_{ijk}}{(z_{j,\max} - z_{j,\min})(m_{k,\max} - m_{k,\min})} \times \delta_{t,t_i}^K \Theta(z - z_{j,\min}) \Theta(z_{j,\max} - z) \Theta(m - m_{k,\min}) \Theta(m_{k,\max} - m), \quad (1)$$

where δ^K denotes Kronecker’s delta and Θ the Heaviside step function. This corresponds to a mathematical description of a three-dimensional histogram, where the probability of finding an object in the bin labelled ijk is f_{ijk} . The type-, redshift- and magnitude-bins are indexed with $i = 1, \dots, N_t$, $j = 1, \dots, N_z$, and $k = 1, \dots, N_m$, respectively. The redshift bins have bounds $(z_{j,\min}, z_{j,\max})$ and the magnitude bins $(m_{k,\min}, m_{k,\max})$. This model does not formally require the bins to be contiguous or of equal size, but they should not overlap. In practice they should simply tile the redshift and magnitude ranges of interest and we will adopt contiguous, equal-size, bins so that the histogram can be interpreted as a piecewise constant approximation of $p(t, z, m)$, such that

$$f_{ijk} \equiv \int_{z_{j,\min}}^{z_{j,\max}} \int_{m_{k,\min}}^{m_{k,\max}} p(t_i, z, m) dz dm. \quad (2)$$

An example of this 3D histogram representation with $N_t = 3$ and $N_z = N_m = 15$ is shown in Fig. 1. It is constructed from the priors $p(t, z, m)$ implemented in the BPZ code, which are parametrized with $p(t|m) = f_t \exp[-k_t(m - 20)]$ and $p(z|t, m) \propto z^{\alpha_t} \exp\{-[z/z_{mt}(m)]^{\alpha_t}\}$. The parameters f_t , k_t , z_{mt} and α_t are taken from table 1 of Benitez (2000) and correspond to the E/S0, Sbc, and Irr templates from Coleman et al. (1980), shown in Fig. 2. In what follows we will sometimes refer to the type-redshift and redshift distributions (with other parameters marginalized over); these can be written as $p(t, z | \{\sum_k f_{ijk}\})$ and $p(z | \{\sum_{ik} f_{ijk}\})$, respectively.

3 INFERENCE METHODOLOGY

We now turn to the problem of inferring the parameters $\{f_{ijk}\}$ from a set of N_{gal} galaxies, the properties of which are denoted by $\{t_g, z_g, m_g\}$ with $g \in \{1, \dots, N_{\text{gal}}\}$. We first consider the ‘noiseless’ case where these properties are assumed to be known for all objects (Section 3.1) before treating the more general case where the individual galaxies’ properties must be inferred simultaneously with the distributions of interest from noisy photometric data (Section 3.2). In both cases we use a uniform (maximally uninformative) prior on the coefficients $\{f_{ijk}\}$, i.e. subject to the constraints $0 \leq f_{ijk} \leq 1$ for

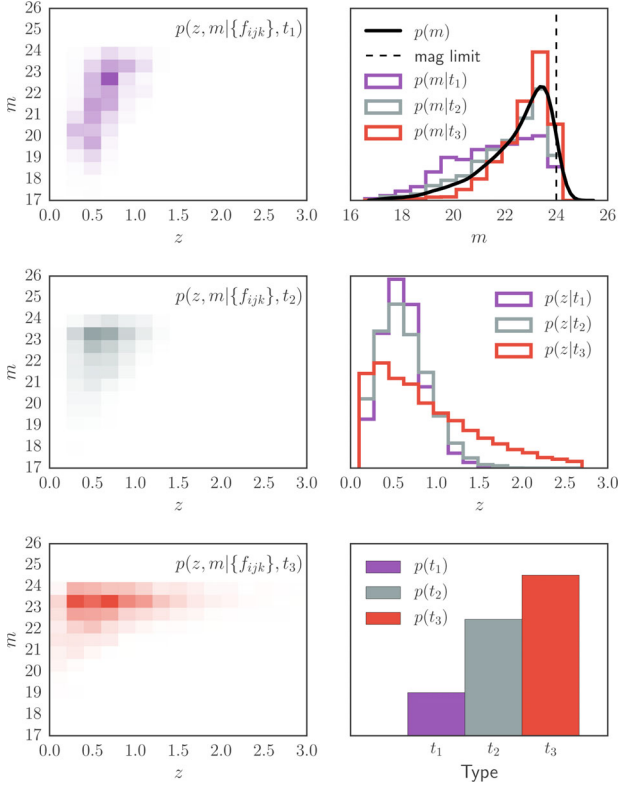


Figure 1. Marginal and conditional distributions describing the binned versions of the target distribution $p(t, z, m)$, parametrized by $\{f_{ijk}\}$. We aim to estimate these distributions using noisy photometric observations.

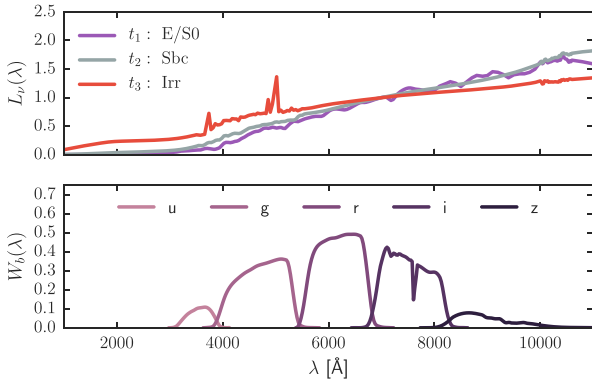


Figure 2. Galaxy spectral energy distributions and filter response curves used in this work. The spectral templates are taken from Coleman et al. (1980) and are normalized to 1 at $\lambda = 7500 \text{ \AA}$. The filters are the SDSS *ugriz* photometric bands (Fukugita et al. 1996).

all ijk and $\sum_{ijk} f_{ijk} = 1$, which is a Dirichlet distribution, given by

$$p(\{f_{ijk}\}) = (N_t N_z N_m - 1)! \delta_D \left(1 - \sum_{ijk} f_{ijk} \right) \times \prod_{i=1}^{N_t} \prod_{j=1}^{N_z} \prod_{k=1}^{N_m} \Theta(f_{ijk}), \quad (3)$$

where $\delta_D(x)$ is the Dirac delta function. A summary of our notation is provided in Table 1.

Table 1. Summary of our notation.

z	Redshift
m	Apparent magnitude in the reference band
i	Index of galaxy type t_i
j	Index of redshift bin with bounds $(z_{j,\min}, z_{j,\max})$
k	Index of magnitude bin with bounds $(m_{k,\min}, m_{k,\max})$
N_t	Number of types
N_z	Number of redshift bins
N_m	Number of magnitude bins
g	Index of galaxy
N_{gal}	Total number of galaxies in the sample
n_{ijk}	Galaxy count in the ijk -th type-redshift-magnitude bin
$\{n_{ijk}\}$	Set of all galaxy counts n_{ijk} , summing to N_{gal}
f_{ijk}	Fractional galaxy count in the ijk -th bin
$\{f_{ijk}\}$	Set of all fractional bin counts f_{ijk} , summing to 1
z_g, t_g, m_g	Properties of the g -th galaxy
$\{z_g, t_g, m_g\}$	Set of properties of all galaxies in the sample
$L_{v,t}$	Spectral template of the type t
b	Index of photometric band
N_b	Number of photometric bands
W_b	b -th photometric filter
$F_b(t, z)$	Noiseless photometric flux (type t , redshift z , band b)
$\{\hat{F}_b\}_g$	N_b observed photometric fluxes of the g -th galaxy
$\{\hat{F}_{b,g}\}$	Observed photometric fluxes for all N_{gal} galaxies

3.1 Noiseless case

In the noiseless case a set of sufficient statistics for the full list of galaxy properties, $\{t_g, z_g, m_g\}$, is the numbers in each of the type, redshift and magnitude bins, which is given by

$$n_{ijk} = \sum_g \delta_{t_g, t_i}^K \Theta(z_g - z_{j,\min}) \Theta(z_{j,\max} - z_g) \times \Theta(m_g - m_{k,\min}) \Theta(m_{k,\max} - m_g). \quad (4)$$

The likelihood of the binned data is given by a multinomial distribution,

$$p(\{n_{ijk}\} | \{f_{ijk}\}) = N_{\text{gal}}! \prod_{i=1}^{N_t} \prod_{j=1}^{N_z} \prod_{k=1}^{N_m} \frac{f_{ijk}^{n_{ijk}}}{n_{ijk}!}. \quad (5)$$

Combining the above prior and likelihood leads to a posterior that, like the prior, is a Dirichlet distribution, which reads

$$p(\{f_{ijk}\} | \{n_{ijk}\}) = (N_{\text{gal}} + N_t N_z N_m - 1)! \delta_D \left(1 - \sum_{ijk} f_{ijk} \right) \times \prod_{i=1}^{N_t} \prod_{j=1}^{N_z} \prod_{k=1}^{N_m} \frac{\Theta(f_{ijk}) f_{ijk}^{n_{ijk}}}{n_{ijk}!}. \quad (6)$$

Thus, inferring the parameters $\{f_{ijk}\}$ when the types, redshifts and magnitudes of the sources are known is possible, thanks to the analytic posterior distribution that only requires the bin counts $\{n_{ijk}\}$. Before we map this formalism to photometric galaxy survey observations, we make three important remarks.

First, in the limit of uncorrelated bins and large bin counts $\{n_{ijk}\}$, the marginalized mean and variance on f_{ijk} both reduce to n_{ijk}/N_{gal} , which is the classical approximate histogram estimator. However, this estimator fails in the regime of strong inter-bin correlations or quasi-empty bins. This is illustrated in Appendix A.

Second, the equations of this section imply an important property of the Dirichlet model: it is agnostic to the nature of the bins,

i.e. to their size or physical interpretation. These are only needed for the calculation of the number counts n_{ijk} . This is a manifestation of the categorical nature of the model: objects drawn from N categories follow a multinomial distribution, and the posterior distribution on its parameters is a Dirichlet distribution. In practice, the only consequence of this property is the insensitivity of the model to the nature of the binning or the order of the labelling ijk , as all the previous equations could have been casted with a unique categorical label $\ell \equiv ijk$. The Dirichlet posterior would then read $p(\{f_\ell\}|\{n_\ell\}) \propto \prod_\ell f_\ell^{n_\ell}/n_\ell!$. Yet, the explicit three-dimensional formalism above will prove essential for the clarity of Section 3.2, where types, redshifts and magnitudes are distinct physical quantities in the hierarchical model, and are marginalized over separately.

Third, prior knowledge about the correlations between bins can be incorporated by considering a prior $p(\{f_{ijk}\})$ following a generalized Dirichlet distribution. In this case, the posterior distribution on $\{f_{ijk}\}$ is also a generalized Dirichlet distribution, and the model is no longer agnostic to the nature and order of the bins. We do not consider this case in this paper but we highlight that this extension does not affect the methods presented below. In particular, in the limit of a large number of galaxies or weak priors on the inter-bin correlations the generalized Dirichlet distribution reduces to a standard Dirichlet case.

3.2 Inference using photometric data

In this section we no longer assume that the types, redshifts and magnitudes of the objects are known. These must now be inferred simultaneously with the underlying distributions, using only noisy photometric observations.

3.2.1 Likelihood

In imaging surveys, the main observable¹ is the photon flux measured in N_b bands indexed by $b = 1, \dots, N_b$,

$$F_b(t, z) = \frac{1+z}{4\pi D_L^2(z)} \int_0^\infty \frac{d\nu}{\nu} L_{\nu,t}[\nu(1+z)] W_b(\nu), \quad (7)$$

where $L_{\nu,t}(\nu)$ is the rest-frame luminosity density of the extragalactic source of type t as a function of frequency ν (see e.g. Hogg et al. 2002). $W_b(\nu)$ are functions characterizing the response of the photometric filters. Fig. 2 shows the three spectral templates as well as the photometric filters used in this paper. The templates are the E/S0, Sbc, and Irr templates from Coleman et al. (1980) as packaged by BPZ (Benitez 2000), while the filters are the SDSS *ugriz* photometric bands (Fukugita et al. 1996) including the effects of extinction and airmass.² Magnitudes in the AB system are related to flux densities via

$$m_b(t, z) = -2.5 \log \left[\frac{F_b(t, z)}{\int g_\nu(\nu) W_b(\nu) d\nu/\nu} \right], \quad (8)$$

with $g_\nu(\nu) = 3631 \text{ Jy}$. Finally, we take the reference magnitude m to be the i band magnitude.

¹ Morphological information about sources is also potentially important: it enables numerous cosmological studies and improves the quality of star-galaxy separation and redshift estimation. However, we focus on magnitude information here in order to exploit widely used libraries of spectral templates, which typically do not use morphological information.

² <http://classic.sdss.org/dr7/instruments/imager>

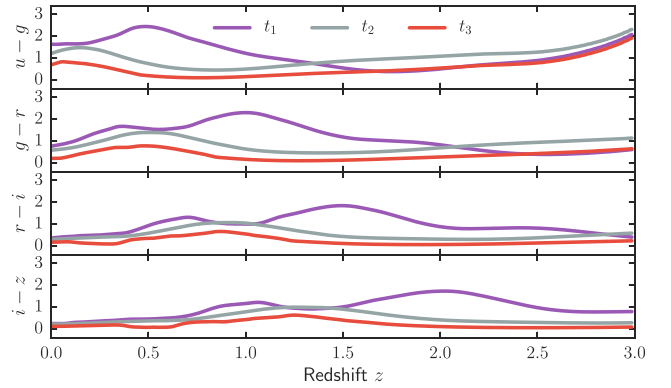


Figure 3. Tracks of the three templates shown in Fig. 2 in the space of SDSS colours as a function of redshift.

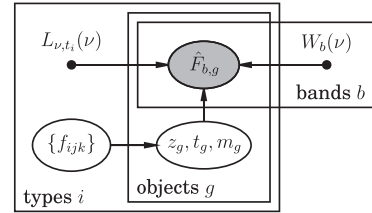


Figure 4. Graphical representation of the hierarchical model, following the notation summarized in Table 1. Dots, circles and shaded circles indicate fixed quantities, parameters to be inferred, and observed quantities, respectively. Arrows express dependence and boxes replication. N_{gal} galaxies with fluxes measured in N_b photometric bands are modelled using a set of N_t spectral templates $L_{\nu,t_i}(\nu)$. The distribution of the intrinsic galaxy properties (type, redshift, and magnitude in the reference band) is parametrized with $\{f_{ijk}\}$.

For a given galaxy g , noisy measurements of its photometric fluxes are available,

$$\{\hat{F}_b\}_g = (\hat{F}_{1,g}, \dots, \hat{F}_{N_b,g}), \quad (9)$$

with errors $\sigma_{\hat{F}_{1,g}}, \dots, \sigma_{\hat{F}_{N_b,g}}$. We define a multidimensional Gaussian likelihood function in flux space,

$$p(\{\hat{F}_b\}_g | t, z, m) = \prod_{b=1}^{N_b} \mathcal{N} \left[F_b(t, z); a \hat{F}_{b,g}, \sigma_{\hat{F}_{b,g}}^2 \right], \quad (10)$$

where $\mathcal{N}(x; \mu, \sigma^2)$ denotes the standard normal distribution of mean μ and variance σ^2 , evaluated at x . Note that this likelihood does not depend on m , and that a is the (arbitrary) template normalization, which we marginalize over analytically following Benitez (2000). A different form could be adopted (e.g. in colour space, provided the correlations between colours sharing a common band was incorporated), with no impact on the methods and conclusions of this paper.

As each spectral template $L_{\nu,t}$ is redshifted and integrated in the photometric bands, it forms a set of tracks in flux–redshift or colour–redshift space, shown in Fig. 3 for the templates and filters considered here.

3.2.2 Efficient parameter inference

The population model and the flux–redshift likelihood can be combined into the hierarchical model shown in Fig. 4. The observed quantities of this model are the set of $N_b \times N_{\text{gal}}$ fluxes, denoted by $\{\hat{F}_{b,g}\}$. The parameters of interest are the N_{gal} triplets of intrinsic

parameters $\{t_g, z_g, m_g\}$ as well as the population parameters $\{f_{ijk}\}$. The full, joint posterior distribution on these parameters reads

$$p(\{t_g, z_g, m_g\}, \{f_{ijk}\} | \{\hat{F}_{b,g}\}) \propto p(\{f_{ijk}\}) \prod_{g=1}^{N_{\text{gal}}} p(\{\hat{F}_b\}_g | z_g, t_g, m_g) p(z_g, t_g, m_g | \{f_{ijk}\}). \quad (11)$$

Unlike the noise-free case of the previous section, the implied normalized posterior does not have an analytic form.

It is possible to directly draw samples from this posterior using a two-step Gibbs sampler because the conditional posterior distributions can be easily sampled. At a given Gibbs iteration, a sample of $\{f_{ijk}\}$ is drawn from $p(\{f_{ijk}\} | \{t_g, z_g, m_g\}, \{\hat{F}_{b,g}\})$. The latter follows the Dirichlet model of equation (6), where the number counts $\{n_{ijk}\}$ are calculated from the triplets $\{t_g, z_g, m_g\}$ of the previous Gibbs iteration. Then, $\{t_g, z_g, m_g\}$ are updated using the newly drawn $\{f_{ijk}\}$ by looping over galaxies and updating each triplet (t_g, z_g, m_g) using

$$p(t, z, m | \{f_{ijk}\}, \{\hat{F}_b\}_g) = \sum_{ijk} \frac{f_{ijk} \times p(\{\hat{F}_b\}_g | t, z, m)}{(z_{j,\text{max}} - z_{j,\text{min}})(m_{k,\text{max}} - m_{k,\text{min}})} \times \delta_{t,t_i}^K \times \Theta(z - z_{j,\text{min}}) \Theta(z_{j,\text{max}} - z) \times \Theta(m - m_{k,\text{min}}) \Theta(m_{k,\text{max}} - m). \quad (12)$$

As for a classical Gibbs sampler, alternately drawing $\{f_{ijk}\}$ and $\{t_g, z_g, m_g\}$ from the previous conditional distributions allows one to explore the full joint posterior distribution of interest.

If one is not interested in the properties of individual galaxies $\{t_g, z_g, m_g\}$, but only in their distributions described by $\{f_{ijk}\}$, then a significant speed up can be achieved by binning $p(\{\hat{F}_b\}_g | t, z, m)$ in the second step of the Gibbs sampler, i.e. by using a binned likelihood

$$p(\{\hat{F}_b\}_g | ijk) = \int_{z_{j,\text{min}}}^{z_{j,\text{max}}} \int_{m_{k,\text{min}}}^{m_{k,\text{max}}} p(\{\hat{F}_b\}_g | t_i, z, m) dm dz. \quad (13)$$

This is because the Dirichlet model only requires the bin counts $\{n_{ijk}\}$, and those can be updated directly by drawing bin locations ijk from the binned likelihood $p(\{\hat{F}_b\}_g | ijk)$, bypassing the parameters $\{t_g, z_g, m_g\}$.

4 SELECTION EFFECTS, REDSHIFT TOMOGRAPHY

4.1 Background

One interesting extension of the formalism presented above is the inclusion of selection effects. An example common to all photometric surveys is the existence of a detection probability characterizing how objects have been added to the sample (including the completeness of the survey) for which the redshift distribution must be estimated. In the example presented in Section 5, we will include this effect via a magnitude limit in the reference band (magnitude limits are not applied to the other bands). This is directly captured by our formalism because the reference magnitude itself is a dimension of the distribution $p(t, z, m)$ and thus parametrized by $\{f_{ijk}\}$. However, any selection effects not directly involving type, redshift or reference magnitude will not be captured in the previous formalism.

A concrete example is the introduction of a second magnitude limit in another band, which would occur if galaxies had to be simultaneously detected in two bands in order to be retained in the sample. This selection effect would naturally get imprinted in the redshift distributions in a non-trivial way (correlated with type, redshift and reference magnitude). Another common example is the splitting of a main galaxy catalogue into redshift bins, which is of interest to study the redshift evolution of galaxy clustering or cosmic shear observables (e.g. Benjamin et al. 2013; Bonnett et al. 2015; Crocce et al. 2015; Leistedt et al. 2015). In this section we show how to deal with such survey-related selection effects.

4.2 General framework

In our hierarchical model, an extra parameter per object must be introduced to support selection effects in general. We denote this parameter by s , and without loss of generality assume that it depends on the observed fluxes $\{\hat{F}_b\}$. Our distributions and observations are now conditioned on s , and the full posterior distribution is

$$p(\{t_g, z_g, m_g\}, \{f_{ijk}\} | \{\hat{F}_{b,g}\}, \{s_g\}) \propto \prod_{g=1}^{N_{\text{gal}}} p(\{\hat{F}_b\}_g | z_g, t_g, m_g, s_g) p(z_g, t_g, m_g | \{f_{ijk}\}, s_g) p(\{f_{ijk}\} | s_g). \quad (14)$$

The motivation for adopting this form is the following: s indicates whether or not an object is included in the sample of interest. All objects in the sample have the same value of s . Therefore, we are truly interested in inferring the parameters $\{f_{ijk}\}$ describing $p(t, z, m | \{f_{ijk}\}, s)$, which is modelled as a piecewise function as above. In other words, the formalism of the previous sections can be used to estimate $\{f_{ijk}\}$ with no other changes than using the modified likelihood function $p(\{\hat{F}_b\}_g | z, t, m, s)$. The latter can be expressed in terms of our original likelihood via

$$p(\{\hat{F}_b\}_g | s, t, z, m) = \frac{p(\{\hat{F}_b\}_g | t, z, m) p(s | \{\hat{F}_b\}_g)}{p(s | t, z, m)}, \quad (15)$$

where we have omitted the g subscripts for conciseness. The first term of the numerator is the original flux–redshift likelihood, while the second term describes the selection effect under consideration. The denominator can be expressed as a marginalization over possible fluxes of these two terms,

$$p(s | t, z, m) = \int p(s | \{\hat{F}_b\}_g) p(\{\hat{F}_b\}_g | t, z, m) d\hat{F}_1 \dots d\hat{F}_{N_b}. \quad (16)$$

Carrying out the inference with $p(\{\hat{F}_b\}_g | s, t, z, m)$ instead of $p(\{\hat{F}_b\}_g | t, z, m)$ will produce posterior distributions on the parameters $\{f_{ijk}\}$ that describe $p(t, z, m | \{f_{ijk}\}, s)$, the distribution of interest with the selection effect included.

4.3 Illustration

In order to understand the effect on our likelihood and gain intuition we now consider a simplified example. Let us assume that instead of observing fluxes $\{\hat{F}_b\}$, we were directly observing a noisy redshift estimate \hat{z} with error function $p(\hat{z} | z) = \mathcal{N}(z; \hat{z}, \sigma_z^2)$.

Further, our selection effect is to retain objects for which \hat{z} is in some range $[z_{\text{min}}, z_{\text{max}}]$. In other words we have the top hat function

$$p(s | \hat{z}) = \Theta(\hat{z} - z_{\text{min}}) \Theta(z_{\text{max}} - \hat{z}). \quad (17)$$

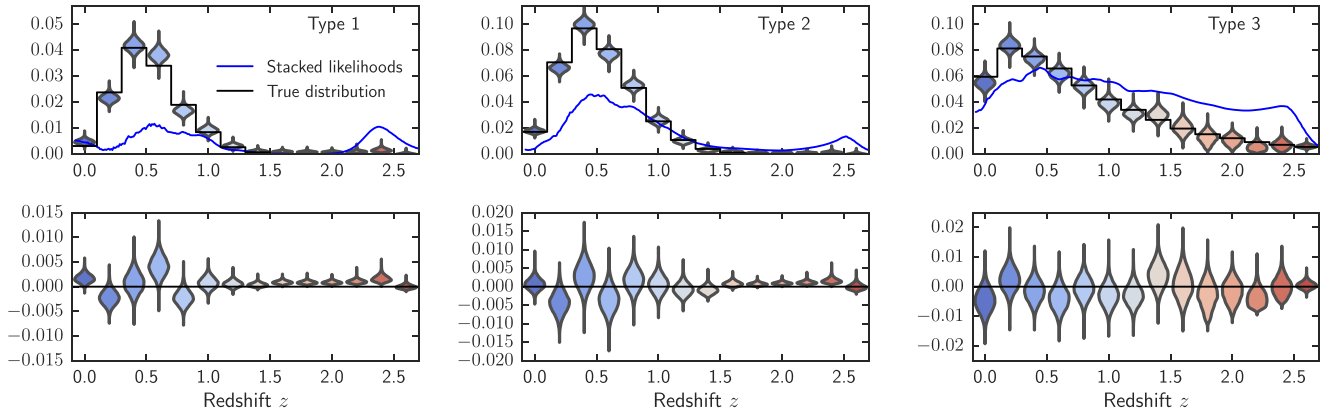


Figure 5. Type-redshift distributions with marginalization over the reference magnitude. The solid lines show the input true distributions while the distributions obtained with the inferred parameters f_{ijk} are shown as violin plots, i.e. box plots whose profiles represent the distribution of parameter values. The bottom panels show the difference and demonstrate that the model not only recovers the input distributions but also provides meaningful error bars. This is illustrated in further detail in Fig. 6.

Using the previous expressions, our likelihood function conditioned on s reads

$$p(\hat{z}|s, z) = \frac{p(s|\hat{z}) p(\hat{z}|z)}{p(s|z)}, \quad (18)$$

with

$$p(s|z) = \int p(s|\hat{z}) p(\hat{z}|z) d\hat{z}. \quad (19)$$

Importantly, because $\mathcal{N}(z; \hat{z}, \sigma_z^2) = \mathcal{N}(z - \hat{z}; 0, \sigma_z^2)$, we can also write $p(s|z)$ as a simple convolution of the selection effect with our error function. In our case this is a top hat smoothed with a Gaussian error function.

Returning to working with fluxes $\{\hat{F}_b\}$, since our likelihood is a multidimensional Gaussian the distribution $p(s|t, z, m)$ can also be seen as a multidimensional convolution. The main difference is that $p(s|\{\hat{F}_b\})$ may take a more complicated form than a top hat. In particular, the case of redshift tomography can be implemented by selecting galaxies whose maximum likelihood redshift estimate, denoted by $z_{\text{ML}}(\{\hat{F}_b\})$, is in a range $[z_{\text{min}}, z_{\text{max}}]$. In this case we have

$$p(s|\{\hat{F}_b\}) = \Theta[z_{\text{ML}}(\{\hat{F}_b\}) - z_{\text{min}}] \Theta[z_{\text{max}} - z_{\text{ML}}(\{\hat{F}_b\})]. \quad (20)$$

As a result, $p(s|t, z, m)$ is a multidimensional convolution of the Gaussian likelihood function with a complicated domain defined by $p(s|\{\hat{F}_b\})$. However, depending on the form of the selection effect and the likelihood function, it might also be possible to draw from $p(\{\hat{F}_b\}|s, t, z, m)$ directly. Both $p(s|t, z, m)$ and $p(\{\hat{F}_b\}|s, t, z, m)$ can be pre-computed for each object prior to inferring the distribution parameters $\{f_{ijk}\}$.

5 DEMONSTRATION ON SIMULATIONS

We now present an illustration of our methodology on simulations of photometric data. Because they do not affect the features of our method, we adopt simple galaxy and noise distributions and do not include selection effects other than a single magnitude limit in the reference band.

We start by drawing $N_{\text{gal}} = 10^4$ i -band magnitudes from a realistic magnitude distribution $p(m)$ with a 5σ magnitude limit of 24, which is shown in the top right panel of Fig. 1. This is achieved by drawing a large number of objects according to $p(m) \propto \exp(m)$, then adding Gaussian noise and only retaining the first N_{gal} objects detected at

more than 5σ significance. In this work we use the noise law of Rykoff, Rozo & Keisler (2015) for all bands. The latter reads

$$\sigma(m) = \frac{2.5}{\ln 10} \sqrt{\left(1 + \frac{F_{\text{noise}}}{F(m)}\right) \frac{1}{F(m)t_{\text{eff}}}}, \quad (21)$$

with $F_{\text{noise}} = t_{\text{eff}} F^2(m_{\text{lim}})/25 - F(m_{\text{lim}})$ and $t_{\text{eff}} = \exp(a + b(m_{\text{lim}} - 21))$. We take $a = 4.56$ and $b = 1$. $F(m)$ simply denotes a conversion from magnitude to fluxes. We then draw the types and redshifts from the multidimensional histogram shown in the other panels of Fig. 1. As emphasized previously, the latter is a binned version of the continuous prior distributions implemented in BPZ. For each object, we use the spectral template and the filters from Fig. 2 to compute the other four magnitudes $ugrz$, and also add noise. These steps provide us with a set of noisy photometric fluxes following the distributions of interest. We now demonstrate that using the method presented above we recover the input distributions as well as the individual redshifts and types of galaxies in this simulated photometric catalogue.

We run the Gibbs sampler³ and obtain samples from the posterior distribution $p(\{t_g, z_g, m_g\}, \{f_{ijk}\} | \{\hat{F}_{b,g}\})$. Drawing 10^4 samples takes about 10 min using a laptop equipped with a 2.8 GHz Intel Core i7 processor.

Fig. 5 shows the distributions of type and redshift (with the reference magnitude marginalized over) obtained with samples $\{f_{ijk}\}$ from the full posterior distribution. For comparison, the stacked likelihood functions of all objects (summed over types and magnitudes) are also shown. The redshift distributions of the three types are successfully recovered with meaningful uncertainties. As observed in this figure and confirmed by the left-hand panels of Fig. 7, the likelihoods distributions are broad and multimodal, demonstrating that stacking them does not yield a meaningful estimate of the distributions of interest. This is a well-known drawback of template-fitting photometric redshift estimation methods: strong distribution priors $p(z, t, m)$ are usually required to deliver reliable redshift estimates (see e.g. Hildebrandt et al. 2012; Benjamin et al. 2013; Bonnett et al. 2015). This is naturally alleviated in our Bayesian hierarchical model, and the mechanism by which the secondary

³ The sampler is initialized at a raw histogram of the maximum likelihood redshifts of the individual objects. In all our tests the sampler converged very quickly and the initialization did not impact the final results.

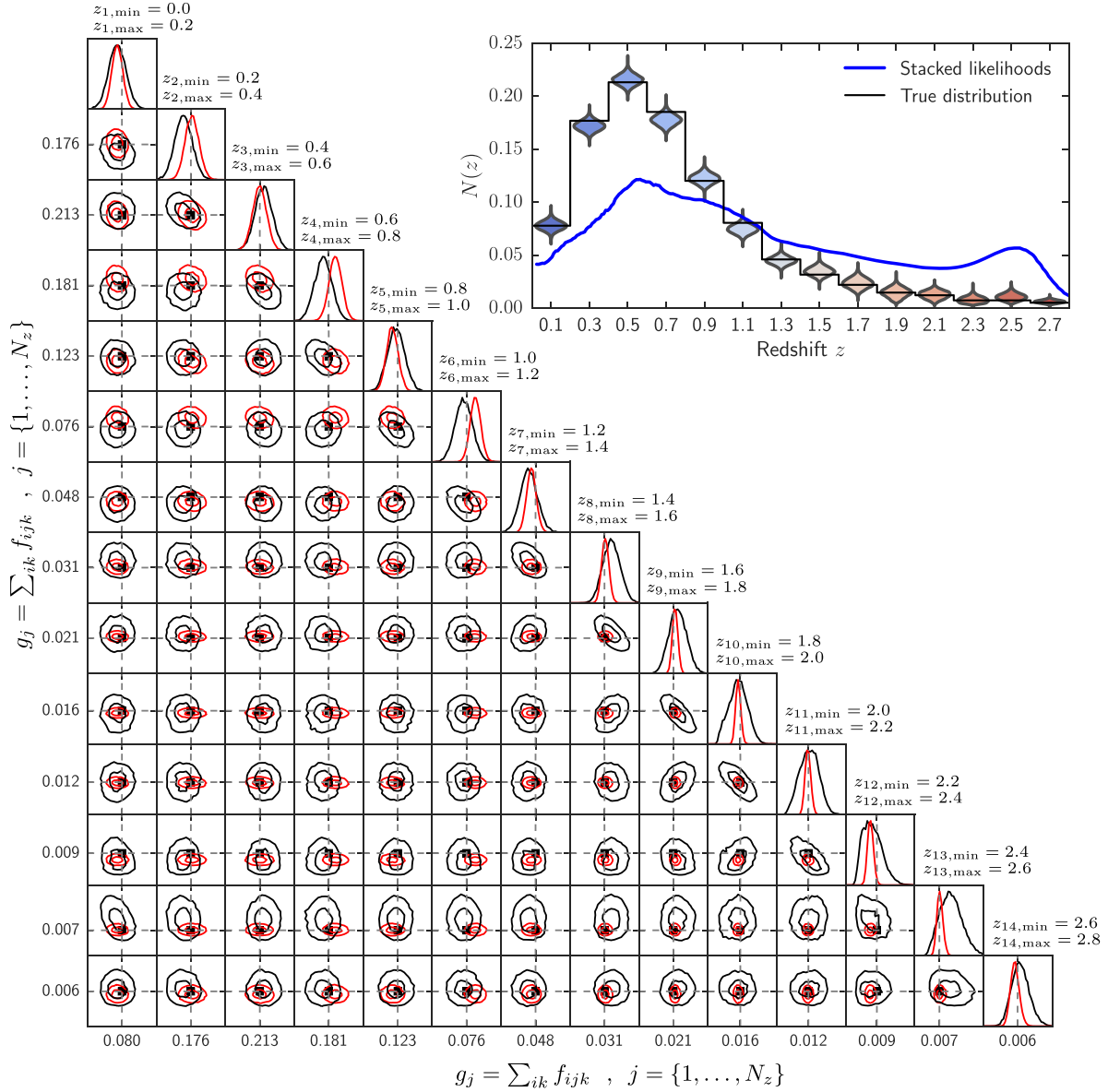


Figure 6. Posterior distributions on the fractional count parameters $g_j = \sum_{ik} f_{ijk}$, i.e. on the type-redshift-magnitude parameters f_{ijk} with marginalization over type and magnitude. The top right panel shows the resulting redshift distributions, with the violin plots illustrating how parameter uncertainties propagate into the redshift distribution estimates. Adjacent bins are significantly anti-correlated, as expected from a Dirichlet model properly harnessing the individual likelihoods. The red contours show the posterior distributions on the same parameters from a noise-free case, i.e. a standard Dirichlet model where the types, redshifts and magnitudes of all objects are known.

peaks are suppressed is illustrated in a simplified setting in Appendix B.

Fig. 6 offers a closer look at the posterior distributions on the overall fractional counts in the redshift bins of the model, i.e. on the parameters $g_j = \sum_{ik} f_{ijk}$ which describe the overall redshift distribution of the sample. As expected, adjacent bins are significantly (anti)correlated, and correlation strength decreases with the separation between bins. Anti-correlation is a feature of the Dirichlet model but is significantly enhanced in our model due to the large support of the individual likelihoods. For comparison, the red contours show a noise-free inference, i.e. samples from a simple Dirichlet model where the number counts are calculated from the true redshifts, types and magnitudes of our simulated catalogue.

The contours are much smaller and exhibit smaller correlations, supporting the previous argument.

Fig. 7 shows the result of our parameter inference on the redshift and type estimates of three randomly chosen individual objects (one per type). The left-hand panels highlight the typical problem of redshift estimation using photometric data: the likelihood functions are multimodal in both redshift and type. The right-hand panels show the maximum a posteriori distributions on type and redshift resulting from constraining the parameters of our hierarchical model. Some of the degeneracies are removed; for example, secondary lobes at high redshift, which are clearly disfavoured by the inferred distributions shown in Fig. 5, have been removed. This demonstrates that the uncertainties in the individual redshifts and types are

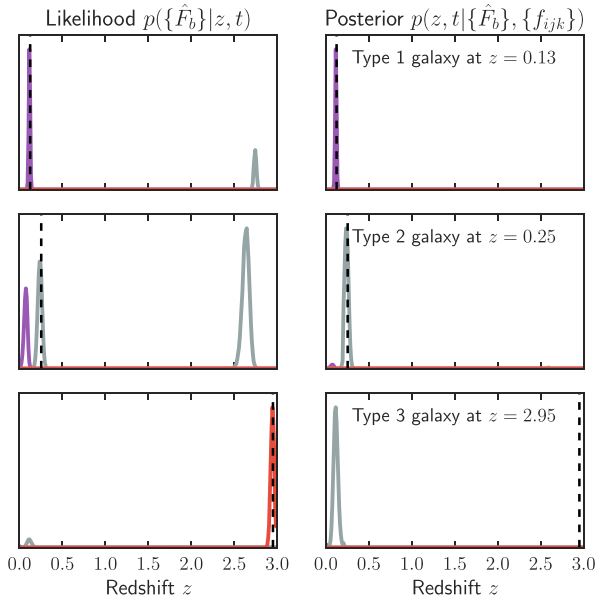


Figure 7. Likelihood functions and posterior distributions for three objects from our simulation – one of each type. The dashed vertical lines show the true redshift. Inferring the individual types and redshifts as well as their underlying population removes some of the strong degeneracies and secondary peaks in the flux–redshift likelihood function.

significantly reduced by including distribution information. Importantly, this new method inferred these redshift distributions directly from the data and propagated their uncertainties into the individual redshift and type estimates. This feature of the hierarchical models is sometimes referred to as ‘Bayesian shrinkage’ as explained in more detail in Appendix B.

6 CONCLUSIONS

Current methods for estimating redshift distributions rely on stacking the posterior probability distributions on the individual redshifts of galaxies. These are typically obtained via template-fitting or machine learning techniques, and are sensitive to the available priors and training sets (e.g. Hildebrandt et al. 2012; Dahlen et al. 2013; Bonnett et al. 2015). The method we presented in this paper is the first to simultaneously infer the redshift distributions and individual redshifts of galaxies from noisy photometric flux measurements.

This new approach can be seen as a generalization of template-fitting methods and has several advantages over existing frameworks. Simultaneously inferring individual redshifts and redshift distributions alleviates the sensitivity to external priors. In addition, the histogram representation of the distributions is very flexible and can represent a range of distributions, capturing features comparable to the bin size. Its relation to the Dirichlet distribution allowed us to design a Gibbs sampling scheme to efficiently infer all the parameters of the model. Any library of spectral templates could be used in this framework to model the flux–redshift relation. In fact, multiple libraries could be used jointly since any degeneracies between the parameters will be inferred correctly, even with highly correlated templates. We have also shown that the inclusion of selection effects such as redshift tomography cuts only affects the likelihood function, not the inference technique.

Our technique provides samples of the posterior distributions on redshift distributions. These samples can be jointly used with galaxy clustering or cosmic shear likelihood functions. Therefore,

our method is the first to consistently exploit spectral templates and provide a way to self-consistently propagate redshift uncertainties in cosmological analyses, which is essential for obtaining meaningful results from ongoing and upcoming photometric surveys.

Our hierarchical model can be extended in various ways to include other physical effects of interest. It could exploit more complicated likelihood functions, for example including combinations of templates and marginalization of extra nuisance parameters such as magnitude zero-points. It could also be interfaced with existing probabilistic approaches for analysing galaxy surveys, such as Bayesian methods for reconstructing the matter density field (e.g. Jasche & Kitaura 2010; Jasche & Wandelt 2013a) and inferring galaxy clustering or cosmic shear power spectra (e.g. Jasche & Wandelt 2013b; Alsing et al. 2016), which do not typically infer redshifts from noisy photometric fluxes. Together, these probabilistic techniques have the potential to unlock the true potential of galaxy surveys by extracting cosmological information while fully accounting for all uncertainties.

ACKNOWLEDGEMENTS

BL thanks Alex I. Maltz and David Hogg for useful discussions. BL and HVP were supported by the European Research Council under the European Community’s Seventh Framework Programme (FP7/2007- 2013)/ERC grant agreement no 306478-CosmicDawn. BL was also supported by the Simons Foundation and DJM by the STFC.

REFERENCES

- Abate A., Newman J. A., Schmidt S. J., 2014, preprint ([arXiv:1410.4498](https://arxiv.org/abs/1410.4498))
- Abbott T. et al., 2005, preprint ([astro-ph/0510346](https://arxiv.org/abs/astro-ph/0510346))
- Abbott T. et al., 2015, preprint ([arXiv:1507.05552](https://arxiv.org/abs/1507.05552))
- Abell P. A. et al., 2009, preprint ([arXiv:0912.0201](https://arxiv.org/abs/0912.0201))
- Albrecht A. et al., 2006, preprint ([astro-ph/0609591](https://arxiv.org/abs/astro-ph/0609591))
- Alsing J., Heavens A., Jaffe A. H., Kiessling A., Wandelt B., Hoffmann T., 2016, *MNRAS*, 455, 4452
- Benítez N., 2000, *ApJ*, 536, 571
- Benjamin J. et al., 2013, *MNRAS*, 431, 1547
- Bonnett C. et al., 2015, preprint ([arXiv:1507.05909](https://arxiv.org/abs/1507.05909))
- Brammer G. B., van Dokkum P. G., Coppi P., 2008, *ApJ*, 686, 1503
- Choi A. et al., 2015, preprint ([arXiv:1512.03626](https://arxiv.org/abs/1512.03626))
- Coleman G. D., Wu C. C., Weedman D. W., 1980, *ApJS*, 43, 393
- Collister A. A., Lahav O., 2004, *PASP*, 116, 345
- Croce M. et al., 2015, preprint ([arXiv:1507.05360](https://arxiv.org/abs/1507.05360))
- Cunha C. E., Huterer D., Busha M. T., Wechsler R. H., 2012, *MNRAS*, 423, 909
- Dahlen T. et al., 2013, *ApJ*, 775, 93
- Feldmann R. et al., 2006, *MNRAS*, 372, 565
- Fukugita M., Ichikawa T., Gunn J. E., Doi M., Shimasaku K., Schneider D. P., 1996, *AJ*, 111, 1748
- Gunn J. E. et al., 2006, *AJ*, 131, 2332
- Heymans C. et al., 2012, *MNRAS*, 427, 146
- Hildebrandt H. et al., 2010, *A&A*, 523, A31
- Hildebrandt H. et al., 2012, *MNRAS*, 421, 2355
- Hogg D. W., Baldry I. K., Blanton M. R., Eisenstein D. J., 2002, preprint ([astro-ph/0210394](https://arxiv.org/abs/astro-ph/0210394))
- Huterer D., Cunha C. E., Fang W., 2013, *MNRAS*, 432, 2945
- Jasche J., Kitaura F. S., 2010, *MNRAS*, 407, 29
- Jasche J., Wandelt B. D., 2012, *MNRAS*, 425, 1042
- Jasche J., Wandelt B. D., 2013a, *MNRAS*, 432, 894
- Jasche J., Wandelt B. D., 2013b, *ApJ*, 779, 15
- Kind M. C., Brunner R. J., 2013, *MNRAS*, 432, 1483
- Leistedt B. et al., 2015, preprint ([arXiv:1507.05647](https://arxiv.org/abs/1507.05647))
- Matthews D. J., Newman J. A., 2010, *ApJ*, 721, 456

Newman J. et al., 2013, preprint ([arXiv:1309.5384](https://arxiv.org/abs/1309.5384))
 Peacock J. A., Schneider P., Efstathiou G., Ellis J. R., Leibundgut B., Lilly S. J., Mellier Y., 2006, preprint ([astro-ph/0610906](https://arxiv.org/abs/astro-ph/0610906))
 Rykoff E. S., Rozo E., Keisler R., 2015, preprint ([arXiv:1509.00870](https://arxiv.org/abs/1509.00870))
 Sadeh I., Abdalla F. B., Lahav O., 2015, preprint ([arXiv:1507.00490](https://arxiv.org/abs/1507.00490))
 Sanchez C. et al., 2014, MNRAS, 445, 1482
 Schmidt S. J., Newman J. A., Abate A., 2014, preprint ([arXiv:1410.4506](https://arxiv.org/abs/1410.4506))
 Weinberg D. H., Mortonson M. J., Eisenstein D. J., Hirata C., Riess A. G., Rozo E., 2013, Phys. Rep., 530, 87

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

zspec_vs_zMAP.png

(<http://mnras.oxfordjournals.org/lookup/suppl/doi:10.1093/mnras/stw1304/-/DC1>).

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

APPENDIX A: GAUSSIAN APPROXIMATION

One of the most common estimators for binned distributions and histograms assumes that the bins are uncorrelated and that the posterior distributions on the fractional count parameters are Gaussian. For the parameters considered in this paper, this can be written as

$$p(\{f_{ijk}\}|\{n_{ijk}\}) = \prod_{ijk} p(f_{ijk}|n_{ijk}) = \prod_{ijk} \mathcal{N}\left(f_{ijk}; \frac{n_{ijk}}{N_{\text{gal}}}, \frac{n_{ijk}}{N_{\text{gal}}}\right). \quad (\text{A1})$$

This form is an approximation of the Dirichlet distribution, which is the correct posterior distribution for histograms. This approximation is accurate in the limit of large, uncorrelated counts n_{ijk} . However, it fails to capture inter-bin correlations. In addition, for low values of n_{ijk} the posterior distribution yields negative values for f_{ijk} , which violates the model assumptions since the fractional counts f_{ijk} must be positive and sum to 1. We highlight these problems using the illustration of Section 5: Fig. A1 shows the posterior distributions on the last two parameters of the second type-redshift distributions of Fig. 5. The correct Dirichlet posterior distributions in both the noiseless and noisy cases are strictly positive. They do not encompass the true values (which are very close to zero) which is a feature of the Dirichlet distribution in this regime. The blue lines of Fig. A1 show the result of approximating the Dirichlet posterior distribution with a multivariate Gaussian distribution with the same mean and diagonal covariance (but no inter-bin correlations). This is similar to what would be obtained by using equation (A1) in place of the Dirichlet distribution in our hierarchical model applied to noisy photometric data. This approximation significantly goes to negative values of f_{ijk} and fails to capture the inter-bin correlations. This highlights the need to employ the correct Dirichlet distribution when inferring the parameters of a binned distribution or a histogram.

APPENDIX B: SUPPRESSION OF SECONDARY REDSHIFT PEAKS

One important aspect of the method presented here for inferring redshift distributions is that the secondary peaks in the likelihood of

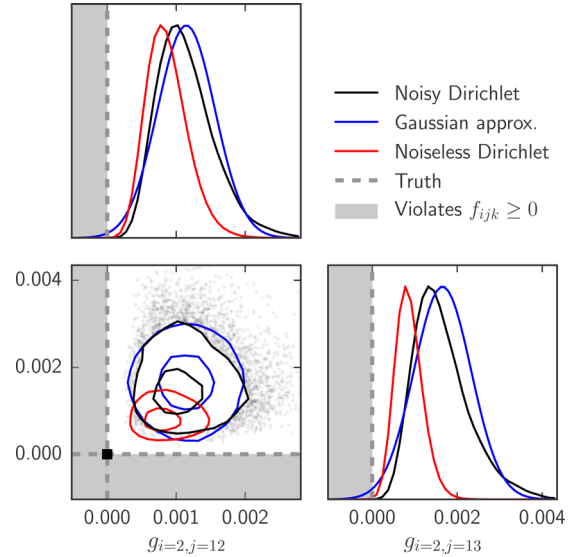


Figure A1. Posterior distributions on the parameters $g_{ij} = \sum_k f_{ijk}$ describing the second type-redshift distributions shown in Fig. 5.

some galaxies are strongly suppressed in the final inferred redshift distribution. This is seen most clearly in the left-hand panel of Fig. 5: the peak in the stacked likelihoods at $z \simeq 2.4$ is completely absent in the redshift distribution as obtained from the Bayesian hierarchical model. The complete absence of any peak associated with these high likelihoods is perhaps counterintuitive – a model with two differently weighted redshift peaks would surely be reasonable? – but the process by which these peaks are suppressed can be seen explicitly by considering a greatly simplified version of the full problem.

The aim, as in Section 2, is to infer the redshift distribution, $p(z|\text{galaxy})$ of a population of galaxies, but here only from a restricted sample of galaxies with strongly bimodal likelihoods. Following equation (1), the redshift distribution is modelled as being piece-wise constant, so that

$$p(z|\text{galaxy}) = \sum_{j=1}^M \frac{\Theta(z - z_{j,\min}) \Theta(z_{j,\max} - z)}{z_{j,\max} - z_{j,\min}} f_j, \quad (\text{B1})$$

where $z_{j+1,\min} = z_{j,\max}$ for $1 \leq j \leq M-1$ (i.e. the bins are contiguous) and f_j is the fraction of galaxies in the j th bin.

The data consist of a sample of N_{gal} very similar galaxies, indexed by $g \in \{1, 2, \dots, N_{\text{gal}}\}$, with photometric data d_g that in all cases leads to a bimodal likelihood of the form

$$p(d_g|z_g) = L_A \delta_D(z - z_A) + L_B \delta_D(z - z_B), \quad (\text{B2})$$

where L_A , L_B , z_A and z_B all determined by d_g . In other words, the photometric data for all of these N_{gal} galaxies are consistent with one of two redshifts, much as is the case for the galaxy whose likelihood is plotted in the top-left panel of Fig. 7.

These galaxies can only contribute to the two bins j_A (for which $z_{j_A,\min} \leq z_A \leq z_{j_A,\max}$) and j_B (for which $z_{j_B,\min} \leq z_B \leq z_{j_B,\max}$); it is assumed that the two redshifts are sufficiently separated that $j_A \neq j_B$. Hence in analysing this sample of N_{gal} galaxies f_j will be zero for all the other bins and so it is only f_{j_A} and f_{j_B} which need to be inferred here. Moreover, $f_{j_B} = 1 - f_{j_A}$, so that f_{j_A} is the only independent population-level parameter.

Taking the prior on f_{j_A} as uniform between 0 and 1 (a special case of the Dirichlet distribution used as the prior in equation 3),

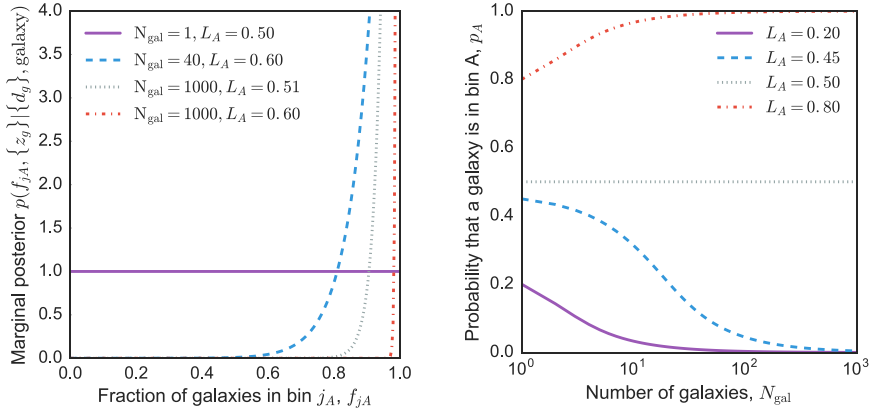


Figure B1. Left: posterior distribution on f_{j_A} given in equation (B5) for various values of N_{gal} and L_A (with $f_{j_B} = 1 - f_{j_A}$ and $L_B = 1 - L_A$). Right: value of p_A , the probability that a given galaxy is at z_A . These highlight how small deviations from $L_A = L_B$ in the likelihood of individual galaxies are amplified in the posterior distributions on p_A and p_B and on the population parameters f_{j_A} and f_{j_B} . This is an illustration of the ‘(Bayesian) shrinkage’ that occurs in the context of hierarchical models.

the joint prior on f_{j_A} and the N_{gal} redshifts $\{z_g\}$ of the galaxies is

$$p(f_{j_A}, \{z_g\} | \text{galaxy}) = \Theta(f_{j_A}) \Theta(1 - f_{j_A}) \times \prod_{g=1}^{N_{\text{gal}}} \left[\frac{\Theta(z_g - z_{j_A, \min}) \Theta(z_{j_A, \max} - z_g)}{z_{j_A, \max} - z_{j_A, \min}} f_{j_A} + \frac{\Theta(z_g - z_{j_B, \min}) \Theta(z_{j_B, \max} - z_g)}{z_{j_B, \max} - z_{j_B, \min}} (1 - f_{j_A}) \right]. \quad (\text{B3})$$

Combining this prior with the likelihood given in equation (B2) leads to the full posterior

$$p(f_{j_A}, \{z_g\} | \{d_g\}, \text{galaxy}) = \Theta(f_{j_A}) \Theta(1 - f_{j_A}) (N_{\text{gal}} + 1) \times \frac{\prod_{g=1}^{N_{\text{gal}}} [f_{j_A} L_A \delta_D(z_g - z_A) + (1 - f_{j_A}) L_B \delta_D(z_g - z_B)]}{\sum_{g=0}^{N_{\text{gal}}} L_A^g L_B^{N_{\text{gal}}-g}}. \quad (\text{B4})$$

Integrating the posterior over the N_{gal} redshifts yields the marginal posterior distribution in f_{j_A} as

$$p(f_{j_A} | \{d_g\}, \text{galaxy}) = \Theta(f_{j_A}) \Theta(1 - f_{j_A}) \times \frac{\sum_{g=0}^{N_{\text{gal}}} \frac{(N_{\text{gal}}+1)!}{g!(N_{\text{gal}}-g)!} (f_{j_A} L_A)^g [(1 - f_{j_A}) L_B]^{N_{\text{gal}}-g}}{\sum_{g=0}^{N_{\text{gal}}} L_A^g L_B^{N_{\text{gal}}-g}}. \quad (\text{B5})$$

Some examples of this posterior for different N_{gal} , L_A and L_B are shown in the left-hand panel of Fig. B1. If $L_A = L_B$ this posterior is uniform between $f_{j_A} = 0$ and $f_{j_A} = 1$, which is the same as the prior – the data provide no reason to prefer either option, no matter how many galaxies are in the sample. Otherwise, taking $L_A > L_B$ (without loss of generality, as the labels are arbitrary), the posterior mode is $f_{j_A} = 1$, even with just one galaxy in the sample. The posterior becomes increasingly sharply peaked with increasing N_{gal} even if L_A is only marginally greater than L_B . Provided only that $L_A > L_B$, $\lim_{N_{\text{gal}} \rightarrow \infty} p(f_{j_A} | \{d_g\}, \text{galaxy}) = \delta_D(f_{j_A} - 1)$.

One result of the potentially sharp posterior in f_{j_A} for large N_{gal} is that the posterior distributions of the redshift of the individual galaxies (which are all the same) can be very different from their associated likelihoods. Integrating over f_{j_A} and $N_{\text{gal}} - 1$ galaxy redshifts (and hence assuming that $N_{\text{gal}} \geq 1$) yields the marginal posterior distribution on the redshift of any one of the N_{gal} galaxies as

$$p(z_g | \{d_g\}, \text{galaxy}) = p_A \delta_D(z_g - z_A) + p_B \delta_D(z_g - z_B), \quad (\text{B6})$$

where

$$p_A = \frac{\sum_{g=0}^{N_{\text{gal}}-1} (g+1) L_A^{g+1} L_B^{N_{\text{gal}}-1-g}}{N_{\text{gal}} \sum_{g=0}^{N_{\text{gal}}} L_A^g L_B^{N_{\text{gal}}-g}} \quad (\text{B7})$$

and

$$p_B = \frac{\sum_{g=0}^{N_{\text{gal}}-1} (g+1) L_A^{N_{\text{gal}}-1-g} L_B^{g+1}}{N_{\text{gal}} \sum_{g=0}^{N_{\text{gal}}} L_A^g L_B^{N_{\text{gal}}-g}} \quad (\text{B8})$$

are, respectively, the probabilities that the galaxy is at z_A and z_B . In the case that $N_{\text{gal}} = 1$ this result reduces to $p_A = L_A/(L_A + L_B)$ and $p_B = L_B/(L_A + L_B)$, and if $L_A = L_B$ then $p_A = p_B = 1/2$. Otherwise (and once again adopting $L_A > L_B$), p_A increases with L_A and N_{gal} , as illustrated in the right-hand panel of Fig. B1. Of particular interest is the case that L_A is only a little higher than L_B but N_{gal} is high: the galaxy is almost certainly at z_B , despite the observed data on this particular object being ambiguous. This is an example of ‘borrowing strength’ or ‘(Bayesian) shrinkage’, in which most of the information about an individual object comes from the population of which it is a member.

This paper has been typeset from a \LaTeX file prepared by the author.